Package ‘dynamicTreeCut’

February 18, 2008

Version 1.10-02

Date 2007-02-18

Title Dynamic Tree Cut Methods for detection of clusters in hierarchical clustering dendrograms.

Author Peter Langfelder <Peter.Langfelder@gmail.com> and Bin Zhang <binzhang.ucla@gmail.com>, with contributions from Steve Horvath <SHorvath@mednet.ucla.edu>

Maintainer Peter Langfelder <Peter.Langfelder@gmail.com>

Depends R, stats

ZipData no

License GPL version 2 or newer

Description Contains methods for detection of clusters in hierarchical clustering dendrograms.

URL http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/

R topics documented:
cutreeDynamic .................................................. 2
cutreeDynamicTree ............................................. 4
cutreeHybrid ..................................................... 5
dynamicTreeCut-package ......................................... 7
indentSpaces .................................................... 8
printFlush ....................................................... 9

Index 10
cutreeDynamic

Adaptive branch pruning of hierarchical clustering dendrograms.

Description

This wrapper provides a common access point for two methods of adaptive branch pruning of hierarchical clustering dendrograms.

Usage

cutreeDynamic(dendro, cutHeight = NULL, minClusterSize = 20, method = "hybrid", distM = NULL, deepSplit = (ifelse(method="dynamic", 1, FALSE)), maxCoreScatter = NULL, minGap = NULL, maxAbsCoreScatter = NULL, minAbsGap = NULL, clusterTrim = 0, labelUnlabeled = TRUE, useMedoids = FALSE, maxDistToLabel = cutHeight, respectSmallClusters = TRUE, verbose = 2, indent = 0)

Arguments

dendro a hierarchical clustering dendrogram such as one returned by hclust.
cutHeight Maximum joining heights that will be considered.
minClusterSize Minimum cluster size.
method Chooses the method to use. Recognized values are "hybrid" and "tree".
distM Only used for method "hybrid". The distance matrix used as input to hclust. If not given and method == "hybrid", the function will issue a warning and default to method = "tree".
depthSplit For method "hybrid", can be either logical or integer in the range 0 to 3. For method "tree", must be logical. In both cases, provides a rough control over sensitivity to cluster splitting. The higher the value (or if TRUE), the more and smaller clusters will be produced. For the "hybrid" method, a finer control can be achieved via maxCoreScatter and minGap below.
maxCoreScatter Only used for method "hybrid". Maximum scatter of the core for a branch to be a cluster, given as the fraction of cutHeight relative to the 5th percentile of joining heights. See Details.
minGap Only used for method "hybrid". Minimum cluster gap given as the fraction of the difference between cutHeight and the 5th percentile of joining heights.
maxAbsCoreScatter Only used for method "hybrid". Maximum scatter of the core for a branch to be a cluster given as absolute heights. If given, overrides maxCoreScatter.
\textit{cutreeDynamic}

- \textbf{minAbsGap} Only used for method "hybrid". Minimum cluster gap given as absolute height difference. If given, overrides \texttt{minGap}.

- \textbf{clusterTrim} Only used for method "hybrid". Specifies the fraction of the cluster height that will be trimmed before stage 2; all objects joining above the 1-\texttt{ClusterTrim} fraction of the cluster height will be trimmed. If set to 0 or less, nothing will be trimmed.

- \textbf{labelUnlabeled} Only used for method "hybrid". If \texttt{TRUE}, the second (PAM-like) stage will be performed.

- \textbf{useMedoids} Only used for method "hybrid" and only if \texttt{labelUnlabeled==TRUE}. If \texttt{TRUE}, the second stage will use object to medoid distance; if \texttt{FALSE}, it will use average object to cluster distance. The default (\texttt{FALSE}) is recommended.

- \textbf{maxDistToLabel} Only used for method "hybrid" and only if \texttt{labelUnlabeled==TRUE}. Maximum object distance to closest cluster that will result in the object assigned to that cluster.

- \textbf{respectSmallClusters} Only used for method "hybrid" and only if \texttt{labelUnlabeled==TRUE}. If \texttt{TRUE}, branches that failed to be clusters in stage 1 only because of insufficient size will be assigned together in stage 2. If \texttt{FALSE}, all objects will be assigned individually.

- \textbf{verbose} Controls the verbosity of the output. 0 will make the function completely quiet, values up to 4 gradually increase verbosity.

- \textbf{indent} Controls indentation of printed messages (see \texttt{verbose} above). Each unit adds two spaces before printed messages; useful when several functions’ output is to be nested.

\section*{Details}

This is a wrapper for two related but different methods for cluster detection in hierarchical clustering dendrograms.

In order to make the shape parameters \texttt{maxCoreScatter} and \texttt{minGap} more universal, their values are interpreted relative to \texttt{cutHeight} and the 5th percentile of the merging heights (we arbitrarily chose the 5th percentile rather than the minimum for reasons of stability). Thus, the absolute maximum allowable core scatter is calculated as \texttt{maxCoreScatter} * (\texttt{cutHeight} - \texttt{refHeight}) + \texttt{refHeight} and the absolute minimum allowable gap as \texttt{minGap} * (\texttt{cutHeight} - \texttt{refHeight}), where \texttt{refHeight} is the 5th percentile of the merging heights.

\section*{Value}

A vector of numerical labels giving assignment of objects to modules. Unassigned objects are labeled 0, the largest module has label 1, next largest 2 etc.

\section*{Author(s)}

Peter Langfelder, \texttt{(Peter.Langfelder@gmail.com)}
References

See Also
hclust, cutreeHybrid, cutreeDynamicTree.

cutreeDynamicTree  Dynamic dendrogram pruning based on dendrogram only

Description
Detect clusters in a hierarchical dendrogram using a variable cut height approach. Uses only the information in the dendrogram itself is used (which may give incorrect assignment for outlying objects).

Usage
cutreeDynamicTree(dendro, maxTreeHeight = 1, deepSplit = TRUE, minModuleSize = 50)

Arguments
dendro Hierarchical clustering dendrogram such produced by hclust.
maxTreeHeight Maximum joining height of objects to be considered part of clusters.
deepSplit If TRUE, method will favor sensitivity and produce more smaller clusters. When FALSE, there will be fewer bigger clusters.
minModuleSize Minimum module size. Branches containing fewer than minModuleSize objects will be left unlabeled.

Details
A variable height branch pruning technique for dendrograms produced by hierarchical clustering. Initially, branches are cut off at the height maxTreeHeight; the resulting clusters are then examined for substructure and if subclusters are detected, they are assigned separate labels. Subclusters are detected by structure and are required to have a minimum of minModuleSize objects on them to be assigned a separate label. A rough degree of control over what it means to be a subcluster is implemented by the parameter deepSplit.

Value
A vector of numerical labels giving assignment of objects to modules. Unassigned objects are labeled 0, the largest module has label 1, next largest 2 etc.
cutreeHybrid

Author(s)
Bin Zhang, ⟨binzhang.ucla@gmail.com⟩, with contributions by Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩.

References
http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting

See Also
cutreeHybrid

cutreeHybrid

Hybrid adaptive tree cut for hierarchical clustering dendrograms.

Description
Detect clusters in a dendrogram produced by the function hclust.

Usage
cutreeHybrid(dendro, distM, cutHeight = NULL, minClusterSize = 20,
  deepSplit = 1,
  maxCoreScatter = NULL, minGap = NULL,
  maxAbsCoreScatter = NULL, minAbsGap = NULL, clusterTrim = 0,
  labelUnlabeled = TRUE,
  useMedoids = FALSE, maxDistToLabel = cutHeight,
  respectSmallClusters = TRUE, verbose = 2, indent = 0)

Arguments
dendro a hierarchical clustering dendrogram such as one returned by hclust.
distM Distance matrix that was used as input to hclust.
cutHeight Maximum joining heights that will be considered.
minClusterSize Minimum cluster size.
depthSplit Either logical or integer in the range 0 to 3. Provides a rough control over sensitivity to cluster splitting. The higher the value, the more and smaller clusters will be produced. A finer control can be achieved via maxCoreScatter and minGap below.
maxCoreScatter Maximum scatter of the core for a branch to be a cluster, given as the fraction of cutHeight relative to the 5th percentile of joining heights. See Details.
minGap Minimum cluster gap given as the fraction of the difference between cutHeight and the 5th percentile of joining heights.
maxAbsCoreScatter  
Maximum scatter of the core for a branch to be a cluster given as absolute heights. If given, overrides maxCoreScatter.

minAbsGap  
Minimum cluster gap given as absolute height difference. If given, overrides minGap.

clusterTrim  
Specifies the fraction of the cluster height that will be trimmed before stage 2; all objects joining above the \(1 - \text{ClusterTrim}\) fraction of the cluster height will be trimmed. If set to 0 or less, nothing will be trimmed.

labelUnlabeled  
If TRUE, the second (PAM-like) stage will be performed.

useMedoids  
If TRUE, the second stage will be use object to medoid distance; if FALSE, it will use average object to cluster distance. The default (FALSE) is recommended.

maxDistToLabel  
Maximum object distance to closest cluster that will result in the object assigned to that cluster.

respectSmallClusters  
If TRUE, branches that failed to be clusters in stage 1 only because of insufficient size will be assigned together in stage 2. If FALSE, all objects will be assigned individually.

verbose  
Controls the verbosity of the output. 0 will make the function completely quiet, values up to 4 gradually increase verbosity.

indent  
Controls indentation of printed messages (see verbose above). Each unit adds two spaces before printed messages; useful when several functions’ output is to be nested.

Details

The function detects clusters in a hierarchical dendrogram based on the shape of branches on the dendrogram. For details on the method, see [http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting](http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting).

In order to make the shape parameters \(\text{maxCoreScatter}\) and \(\text{minGap}\) more universal, their values are interpreted relative to \(\text{cutHeight}\) and the 5th percentile of the merging heights (we arbitrarily chose the 5th percentile rather than the minimum for reasons of stability). Thus, the absolute maximum allowable core scatter is calculated as \(\text{maxCoreScatter} \times (\text{cutHeight} - \text{refHeight}) + \text{refHeight}\) and the absolute minimum allowable gap as \(\text{minGap} \times (\text{cutHeight} - \text{refHeight})\), where \(\text{refHeight}\) is the 5th percentile of the merging heights.

Value

A list containing the following elements:

- **labels**: Numerical labels of clusters, with 0 meaning unassigned, label 1 labeling the largest cluster etc.
- **cores**: Numerical labels indicating cores of found clusters.
- **smallLabels**: Numerical labels for branches that failed to be recognized clusters only because of insufficient number of objects.
trimmed Numerical labels indicating objects that have been trimmed from clusters.
branches A list detailing the detected branch structure.

Author(s)
Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

References

See Also
hclust, as.dist

dynamicTreeCut-package
Methods for detection of clusters in hierarchical clustering dendrograms.

Description
Contains methods for detection of clusters in hierarchical clustering dendrograms.

Details
Package: dynamicTreeCut
Version: 1.10-02
Date: 2007-02-18
Depends: R, stats
ZipData: no
License: GPL version 2 or newer
URL: http://www.genetics.ucla.edu/labs/horvath/CoexpressionNetwork/BranchCutting/
Packaged: Mon Feb 18 12:32:38 2008; plangfelder
Built: R 2.4.1; ; 2008-02-18 12:48:47; unix

Index:
cutreeDynamic Adaptive branch pruning of hierarchical clustering dendrograms.
cutreeDynamicTree Dynamic dendrogram pruning based on dendrogram only
cutreeHybrid Hybrid adaptive tree cut for hierarchical clustering dendrograms.
indentSpaces Spaces for indented output.
printFlush

treecut-package

Print arguments and flush the console.
Methods for detection of clusters in hierarchical clustering dendrograms.

Author(s)

Peter Langfelder <Peter.Langfelder@gmail.com> and Bin Zhang <binzhang.ucla@gmail.com>, with contributions from Steve Horvath <SHorvath@mednet.ucla.edu>

Maintainer: Peter Langfelder <Peter.Langfelder@gmail.com>

---

indentSpaces is used to create spaces for indented output.

Description

Returns a character string containing two times `indent` spaces.

Usage

```r
indentSpaces(indent = 0)
```

Arguments

- `indent` Desired level of indentation. The number of returned spaces will be twice this argument.

Value

A character string containing spaces, of length twice `indent`.

Author(s)

Peter Langfelder, <Peter.Langfelder@gmail.com>

Examples

```r
spaces = indentSpaces(0);
print(paste(spaces, "This output is not indented..."));
spaces = indentSpaces(1);
print(paste(spaces, "...while this one is."))
```
printFlush  

Print arguments and flush the console.

Description

Passes all its arguments unchanged to the standard `print` function; after the execution of `print` it flushes the console, if possible.

Usage

```
printFlush(...)  
```

Arguments

...  

Arguments to be passed to the standard `print` function.

Details

Passes all its arguments unchanged to the standard `print` function; after the execution of `print` it flushes the console, if possible.

Value

Returns the value of the `print` function.

Author(s)

Peter Langfelder, ⟨Peter.Langfelder@gmail.com⟩

See Also

`print`
Index

*Topic cluster
  cutreeDynamic, 1
  cutreeHybrid, 5
*Topic misc
  cutreeDynamicTree, 4
*Topic package
  dynamicTreeCut-package, 7
*Topic print
  indentSpaces, 8
  printFlush, 8

as.dist, 7

cutreeDynamic, 1
  cutreeDynamicTree, 3, 4
  cutreeHybrid, 3, 5, 5

dynamicTreeCut
  (dynamicTreeCut-package), 7
  dynamicTreeCut-package, 7

hclust, 3, 4, 7

indentSpaces, 8

print, 9
  printFlush, 8